63-4

SP-1120

Coding Clinical Laboratory Data For Automatic Storage and Retrieva

Leonard D. Gross

9 May 1963

(SP Series)

SP-1120

Coding Clinical Laboratory Data For Automatic Storage and Retrieval[1]

by

Leonard D. Gross

May 9, 1963

SYSTEM DEVELOPMENT CORPORATION, SANTA MONICA, CALIFORNIA

This paper was prepared for publication outside SDC.

CODING CLINICAL LABORATORY DATA FOR AUTOMATIC STORAGE AND RETRIEVAL

by

Leonard D. Gross

ABSTRACT

A series of clinical laboratory codes have been developed to accept and store Urinalysis, Blood Chemistry, and Hematology test results for automatic data processing. The codes, although constructed as part of a computerized hospital simulation, have been able to handle the results of every laboratory test that they have encountered. The unique feature of these codes is that they can accept conventionally recorded qualitative as well as quantitative test results. Consequently, clinical test results need not be arbitrarily stratified, standardized, or altered in any way to be coded. This paper describes how the codes were developed and presents a listing of the Urinalysis codes. Five criteria used in developing the codes are outlined and the problem of multiple-synonymous terminology is discussed. A solution to the problem is described. Flexible, computer-produced, composite laboratory reports are also discussed, along with reproduction of such a report. The paper concludes that even though many problems remain unsolved, the next ten years could witness the emergence of a practical automated information system in the laboratory.

INTRODUCTION

A hypothetical hospital, consisting of six wards with ten beds on each ward, has been simulated on an Electronic Data Processing system by the System Development Corporation (SDC), in conjunction with the Veterans Administration (VA) Department of Medicine and Surgery. Urinalysis, Blood Chemistry, and Hematology laboratory test results comprised a substantial portion of the clinical data handled in the simulation. Test results to be processed by the computer were first extracted from clinical records of patients recently discharged from the Wadsworth VA Hospital in Los Angeles. The results were then encoded onto punched cards, input to the computer system, and stored on disc files, to be decoded and displayed upon request. Considerable effort was expended developing codes to transform these test results into an efficient configuration for automatic storage and retrieval. This paper describes how the codes were developed and presents a listing of the Urinalysis Codes.

LABORATORY DATA COLLECTION

There are two basic means of collecting laboratory data for entry into an automatic computing system: one, data can be collected directly in machine-usable code as electrical signals from testing devices; and two, test results can be transformed manually from conventional forms into machine-usable code. Future laboratory input systems will, in the author's opinion, primarily rely on the first method of collecting test results. In such a system, testing devices will be in direct communication with computing apparatus. Results of a laboratory test can then be automatically transmitted to the computer the instant they are available and entered in the patient's electronically-stored clinical record. However, numerous economic and technical problems must be resolved before such devices can be placed in the laboratory.

Therefore, we chose to simulate the second method of collecting laboratory results. The knowledge acquired by manual coding of test data will be extremely useful in the future as automatic input apparatus becomes more readily available. This knowledge also enables us to begin automating many laboratory activities much sooner than if we had to wait for a completely automatic system.

DEVELOPING THE LABORATORY CODES

The initial task in developing the laboratory codes was to select a representative assortment of laboratory report forms for coding purposes. After examining various records used in the VA Hospital, we selected Urinalysis, Blood Chemistry, and Hematology. These three were chosen because, as a group, they present typical coding requirements. In addition, at least one of the sixty tests on the forms was requested for every patient in the hospital.

Next, we studied communication patterns between the ward and laboratory. This study concentrated on existing methods of collection, storage, and retrieval of patient data. Special attention was devoted to defining shortcomings and inconsistencies of the conventional handwritten report.

The subsequent stage of code development was a comprehensive investigation of Urinalysis, Blood Chemistry, and Hematology reports. Approximately two hundred samples of each report were examined from laboratory records at the Wadsworth VA Hospital in Los Angeles. Data entries were carefully reviewed for the following characteristics:

- Number of characters (numerical and alphabetic) required to record a particular test reading
- Degree of precision used to report the test reading
- Units in which the test measurement was expressed (e.g., gm, mm)
- Upper and lower limits between which the reading is expected to vary
- Terms and expressions which are used synonymously
- Non-listed tests which can be recorded on the report form
- Test results which are described in narrative terminology
- Heading information required to identify the patient, his location, tests ordered, and other associated instructions

Initially, we focused on normal, everyday modes of reporting test results. Once familiarity with typical data-recording techniques was achieved, attention was devoted to the special or unusual example. The majority of these variations resulted from the different ways in which individual laboratory technicians expressed seemingly identical test results. For example, in reporting the amount of a particular microscopic particle, some technicians would quantify their observations by writing "1+," while others preferred descriptive terms such as "occasional," "rare," "infrequent," "small number," etc. For our simulation, the problem was resolved by employing the same code to represent a group of terms previously determined to be synonymous. In the example above, the digit "1" in a particular card column was used to code the terms "1+," or "occasional," "rare," etc. Since all synonymous terms were represented by a single code, one term had to be selected as the preferred output expression. In this case, the term chosen was "1+." Laboratory supervisors provided substantial assistance to us in identifying terms which were used synonymously and in selecting preferred output expressions.

As the last stage of code development, we attempted to determine valid limits for accepting or rejecting test results. Defining an "acceptance interval," however, turned out to be an extremely elusive task. The limits of "acceptable" test results are not universal but vary from disease to disease. To illustrate, a white-cell count of 50,000 would not be unusual for a patient suffering from leukemia; this same cell count, however, would be open to question for a patient suffering from a fractured tibia only. Therefore, it was decided to temporarily postpone the study.

CODING CRITERIA

After we completed the detailed examination of Urinalysis, Blood Chemistry, and Hematology test results, we began the actual construction of laboratory codes. Five criteria were used as a guide. As a mnemonic tool, we referred to the criteria as the five C's of effective coding: conciseness, consistency, comprehensiveness, compatibility, and convenience. The first of these five criteria, conciseness, refers to the brevity and compactness of the code; consistency pertains to the uniformity of the code, i.e., like items will be represented by the same code configuration; comprehensiveness concerns the versatility and inclusiveness of the code, i.e., the ability of the code to represent uniquely all variations of the data being coded; compatibility refers to the logical agreement between the given code and other codes the system processes; convenience applies to the ease and facility with which the data is converted into its coded representation.

One requirement that restricted uniform application of the five criteria was that all clinical reports generated by the computer system had to be compatible with their existing manually-prepared counterparts. Consequently, the codes had to be capable of accepting conventionally recorded test results, including both quantitative and qualitative results. Results could not be arbitrarily stratified, standardized, or altered in any way. The codes were designed so that no reorientation would be required either on the part of the laboratory technician recording test results or on the part of the physician receiving a laboratory printout. As a result, computer-produced reports were as inclusive and detailed as conventional manual reports.

THE LABORATORY CODES

Urinalysis codes developed for the simulation are presented in Table I. Hematology and Blood Chemistry codes are available from the author.[2] Table I is divided into five columns: Current Item Heading, Current Data Usage, Data Code, Number of Characters, and Card Columns. Current Item Heading refers to the various items on the laboratory form, i.e., the printed heading on the present form. The heading includes identifying information and test titles. Current Data Usage refers to the manner in which information is now recorded on the laboratory form. Data Code refers to the punch-card code that will represent the particular item.

Numerical test results (listed as X-X under Current Data Usage) are formatted but not coded. Narrative results are coded as well as formatted. Number of Characters refers to the number of characters required to code and store the given item. Card Columns refers to the columns on a standard 80-column punch-card in which the data will be entered. Card columns 1-5 and 76-80 are reserved for card control data.

Whenever equivalent terms are used, the term which appears with an asterisk (*) is the designated output. For example, the terms "small number," "infrequent," "occasional," "rare," and "1+" are used interchangeably in the hospital and, therefore, all five are coded by the digit "1;" but "1+" has been selected as the preferred output and is designated as "*1+."

The codes presented in Table I were developed primarily for a research-oriented simulation of laboratory procedures. They were not intended to be used as part of an operational, on-going system within the laboratory. However, with certain refinements, they could be adapted to operational use in a clinical laboratory, as indicated by their successful application to actual laboratory data. The codes have been able to handle the results of every Urinalysis, Blood Chemistry, and Hematology test that they have encountered.

The unique feature of the codes is that they can accept conventionally recorded laboratory results. Most of the other coding schemes with which the author is familiar depend on artificially structured results. These other techniques often require that laboratory technicians stratify their results into fixed ranges. The technician records only the range in which a given result falls, rather than the actual result. Such codes cause a loss of information and make the test reports less useful for research or other studies. Another common coding technique involves a binary-format scheme. Here, only the presence or absence of the specimen characteristic is noted. The latter method cannot handle gradations in results and, consequently, is even more limited than stratification.

While the aforementioned coding schemes tend to simplify conversion of numerical test results for statistical studies and computer-generated diagnoses, they may simultaneously degrade the information to the point where it no longer satisfies the needs of the physician. Perhaps their most serious shortcoming is that they cannot handle descriptive observations; they are keyed solely to quantitative results. Many laboratory reports, however, contain narrative descriptions; foremost are microscopic examinations. Microscopic tests are often as important as quantitative readings. Therefore, a comprehensive laboratory system must be able to convert qualitative data into machine-usable code and disseminate it to the ward along with quantitative results. As seen in Table I, the laboratory codes developed for the hospital simulation had provision for handling qualitative as well as quantitative results.

COMPUTER-PRODUCED REPORTS

Storing laboratory data within an automated data processing system provides a tremendous potential for improving the dissemination of test results. It also affords an opportunity to furnish data to the ward or the physician in flexible composite presentations rather than in the restricted stereotyped formats currently used. Figure 1 contains a reproduction of such a composite report. The computer-generated Consolidated Urinalysis Report depicted in Figure 1 was produced for our hospital simulation. This printout identifies the patient (Monroe, S. T., #91432), the date (05/05/62), and presents the results of five test series along with pertinent heading information (one report can accommodate up to eight test series). Combining a group of tests on such a form allows the physician to readily evaluate trends as well as individual results.

A printout similar to that shown in Figure 1 could be quickly generated--on demand--by an automated data processing system. Also, the system could consolidate laboratory test results with vital sign readings, or any patient parameter, for that matter, in a single comprehensive record. In addition, a fully automated information system could provide the laboratory with a projected workload schedule based on pending doctors' laboratory orders, as well as prepare the monthly summary of tests performed.

CONCLUSION

Many problems have to be solved and much work remains to be done before large-scale automated data processing becomes a reality in the clinical laboratory of the modern hospital. One of the foremost problems is how to handle the variable terminology used to record clinical test results. Another is the difficulty of capturing test results within the laboratory in machine-usable code. An allied problem is lack of suitable data input devices. To insure their acceptance in the laboratory, input devices must be convenient to operate and must be reasonably moderate in cost. In addition to these difficulties, a wide variety of legal, educational, and psychological-acceptance problems must be resolved. If a sense of urgency can be developed, it is reasonable to assume that the next ten years will see the emergence of automated information systems for the laboratory.

ACKNOWLEDGMENT

I wish to express my gratitude to the following members of the Laboratory Service at the Wadsworth VA Hospital in Los Angeles, California, for their generous assistance in developing the codes presented herein: Dr. B. G. Fishkin, Chief of Laboratory Services; Mr. G. R. Kingsley, Supervisor of the Biochemistry Laboratory; and Mr. M. Fukushima, Supervisor of the Hematology Laboratory.

TABLE I: URINALYSIS REPORT

CURRENT ITEM HEADING	CURRENT DATA USAGE	DATA CODE	NO. OF CHARACTERS	CARD COLUMN(S)
Patient's Unit Number	5-digit number	Same	5	6-10
Patient's Last, First, and Middle Name	Patient's name	Patient's Initials	3	11-13
(None)	Report Serial Number First Urinalysis Second Urinalysis Third-Eighth Urinalysis	 1 2 3-8	1	14
Time of Report	Report Time	Hour	4	15-18
Date of Report	Report Date	Month, Day	4	19-22
Signature	Signature of Authorized Personnel	Initials	3	23-25
Color-Appearance	Color Yellow Amber Straw Bloody Appearance Clear Hazy Turbid	 1 2 3 4 C H T	2	26 27
Reaction	XX.X [1,2]	Same	3	28-30
Specific Gravity	X.XXX	Same	4	31-34
Albumin	Trace 1+ 2+ 3+ 4+	0 1 2 3 4	1	35
Sugar	Trace *1+; or 1/10% *2+; or 1/4% *3+; or 1/2% *4+; or 2% *negative; or 0%	0 1 2 3 4 5	1	36

TABLE I: URINALYSIS REPORT (CONT'D)

CURRENT ITEM HEADING	CURRENT DATA USAGE	DATA CODE	NO. OF CHARAC-TERS	CARD COLUMN(S)
Acetone	Positive 1+ 2+ 3+ 4+	0 1 2 3 4	1	37
Bile	Negative Positive	N P	1	38
Microscopic Examination	Adjectives *Present; or positive *1+, or occasional; or small number; or rare; or infrequent *2+; or few; or some *3+; or many; or number *4+; or great many; or packed	Adjective Code 0 1 2 3 4	NA	NA
	Casts Hyaline Casts Finely Granular Casts Coarsely Granular Casts RBC Casts Hemoglobin Casts Pus Casts Epithelial Casts Fatty Casts Waxy Casts Broad Casts Modified by adjectives listed above	3 A B C D E F G H I J	2 Per entry	39-44
	*Mucous Threads; or Cylindroids Modified by adjectives listed above	Adjective Code[4]	1	45
	*Amorphous; or amorphous material; or amorphous sediment Modified by adjectives listed above	Adjective Code	1	46

TABLE I: URINALYSIS REPORT (CONT'D)

CURRENT ITEM HEADING	CURRENT DATA USAGE	DATA CODE	NO. OF CHARAC-TERS	CARD COLUMN(S)
Microscopic Examination (cont'd.)	Crystals *Amorphous Phosphate Crystals or Phosphate Crystals	K^3	2 per entry	47-54
	Triple Phosphate Crystals	L		
	Calcium Phosphate Crystals	M		
	Calcium Sulphate Crystals	N		
	Calcium Carbonate Crystals	O		
	Calcium Oxalate Crystals	P		
	Uric Acid Crystals	Q		
	Ammonium Borate Crystals	R		
	Sodium Urate Crystals	S		
	Tyrosine Needles Crystals	T		
	Eucine Spheres Crystals	U		
	Cystine Crystals	V		
	Fatty Acids Crystals	W		
	Indigo Crystals	X		
	Hippuric Acid Crystals	Y		
	Modified by adjectives listed above	Adjective Code		
	Bacteria Modified by adjectives listed above	Adjective Code	1	55
	Epithelial Cells		2	56-59
	Renal Epithelial Cells	R		
	*Caudate Epithelial Cells; or Cylindrical Epithelial Cells	C		
	Squamous Epithelial Cells	S		
	Modified by adjectives listed above	Adjective Code		
	Occult Blood Modified by adjectives listed above	Adjective Code	1	60
	Sperm Cells Modified by adjectives listed above	Adjective Code	1	61

TABLE I: URINALYSIS REPORT (CONT'D.)

CURRENT ITEM HEADING	CURRENT DATA USAGE	DATA CODE	NO. OF CHARAC-TERS	CARD COLUMN(S)
Microscopic Examination (cont'd.)	The term RBC/HPF modified by adjectives listed above or RBC/HPF reported as a number such as 0-5 or 1-3	Adjective Code or The numbers recorded on the lab form (without the dash)	2	62 or 62-63
	The term WBC/HPF modified by adjectives listed above or WBC/HPF reported as a number such as 0-5 or 1-3	Adjective Code or The numbers recorded on the lab form (without the dash)	2	64 or 64-65

[1] The decimal point is not included in the input code for numerical readings.

[2] The interpretation of the symbol XX.X can be illustrated by the following: a result recorded as 9.32 is coded "093;" a result recorded as 9 is coded "090;" and a result recorded as 93 is coded "930."

[3] Two columns on the card are allotted for each type of microscopic particle identified. The first column contains the alpha code representing the particle name; the second column contains the numeric adjective code. The entire 6-character field can accommodate a maximum of three entries.

[4] The presence of the microscopic particle is denoted by an adjective code in the card column assigned to the given particle.

FIGURE 1

URINALYSIS CONSOLIDATED LABORATORY REPORT

PATIENT-MONROE ST NUMBER-91432 05/05/62

TEST	FIRST	SECOND	THIRD	FOURTH	FIFTH	SIXTH	SEVENTH	EIGHTH
WARD	C3E	C3E	C3E	C3E	C3E			
PROCEDURE	EMERGENCY	ROUTINE	ROUTINE	ROUTINE	ROUTINE			
REQUEST DATE	05/01/62	05/01/62	05/01/62	05/01/62	05/01/62			
PHYSICIAN	MEE	MEE	MEE	MEE	MEE			
REPORT DATE	05/01	05/02	05/03	05/04	05/05			
SAMPLE DATE	05/01	05/02	05/03	05/04	05/05			
SAMPLE TIME	0330	0615	0630	0645	0620			
LAB SIGNATURE	PTO	RAS	RAS	RAS	RAS			
COLOR	AMBER	AMBER	AMBER	STRAW	STRAW			
APPEARANCE	HAZY	HAZY	CLEAR	CLEAR	CLEAR			
REACTION	4.0	4.0	4.0	5.0	5.0			
SPECIFIC GRAVITY	1.013	1.002	1.007	1.013	1.018			
ALBUMIN	1+	TRACE	NEGATIVE	NEGATIVE	NEGATIVE			
SUGAR	NEGATIVE	NEGATIVE	NEGATIVE	NEGATIVE	NEGATIVE			
ACETONE	NEGATIVE	NEGATIVE	NEGATIVE	NEGATIVE	NEGATIVE			
BILE	NEGATIVE	NEGATIVE	NEGATIVE	NEGATIVE	NEGATIVE			
MICROSCOPIC								
BACTERIA	1+	1+						
WBC/HPF	4+	3+	7-9	1+	1+			
EPITHELIAL RENAL	4+	2+						

FOOTNOTES

1 This paper is based on research in the area of hospital automation performed by the System Development Corporation's Bio-Medical Systems Department and the Veterans Administration Department of Medicine and Surgery. Consistent with the general policies of the System Development Corporation and the Veterans Administration, the results of this research are being made available with the hope that they will materially assist others engaged in these activities.

2 For a complete listing of the codes, interested readers may obtain a copy of document SP-1120/001/00 ("Laboratory Codes Developed for the Simulated Hospital Experiment") by writing to Mr. Leonard Gross, System Development Corporation, 2500 Colorado Avenue, Santa Monica, California.

REFERENCES

1. Blumberg, Mark S. "Hospital Automation Needs and Prospects." An address delivered to the annual meeting of the California Hospital Association, Santa Barbara, California, October 1960.

2. Jonnard, Raymond. "A Random Selection System for Automatic Dynamic Biochemical Analysis." Annals of the New York Academy of Sciences, Vol. 87 (22 July 1960), pp. 669-728.

3. Schenthal, J. E., Sweeney, J. W., and Nettleton, W., Jr. "Clinical Application of Electronic Data Processing Apparatus." Journal of the American Medical Association, Vol. 178, No. 3 (21 October 1961), pp. 267-270.

4. Schwichtenberg, Albert H., Flickinger, D. D., and Lovelace, R. W. "Development and Use of Medical Machine Record Cards in Astronaut Selection." USAF Medical Journal, Vol. 10, No. 11 (November 1959), pp. 1324-1351.

5. Seligson, David. "Automation and Labor-Saving Devices in the Clinical Laboratory." Connecticut Medicine, Vol. 25, No. 7 (1961), pp. 423-424.

UNCLASSIFIED

System Development Corporation,
Santa Monica, California
CODING CLINICAL LABORATORY DATA FOR
AUTOMATIC STORAGE AND RETRIEVAL.
Scientific rept., SP-1120, by L. D.
Gross. 9 May 1963, 13p., 5 refs.,
1 table, 1 fig.
(Contract V1005P-9571)
Unclassified report

DESCRIPTORS: Information Retrieval.
Data Processing Systems. Medical
Research.

Reports that a series of clinical
laboratory codes have been developed
to accept and store Urinalysis,
Blood Chemistry, and Hematology

UNCLASSIFIED

---

test results for automatic data
processing. Describes how the
codes were developed and presents
a listing of the Urinalysis codes.
Outlines five criteria used in
developing the codes. Discusses
multiple-synonymous terminology
and describes a solution to the
problem. Discusses flexible,
computer-produced, composite labora-
tory reports, and reproduction of such
reports. Concludes that even though
many problems remain unsolved, the
next ten years could witness the
emergence of a practical automated
information system in the laboratory.
See also SP-1120/001/00.

UNCLASSIFIED

UNCLASSIFIED